



THE UNIVERSITY *of* EDINBURGH

Edinburgh Research Explorer

## Colombian Women's Life Patterns: A Multivariate Density Regression Approach

### Citation for published version:

Wade, SK, Piccarreta, R, Cremaschi, A & Antoniano-Villalobos, I 2021, 'Colombian Women's Life Patterns: A Multivariate Density Regression Approach', *Bayesian analysis*. <https://doi.org/10.1214/20-BA1256>

### Digital Object Identifier (DOI):

[10.1214/20-BA1256](https://doi.org/10.1214/20-BA1256)

### Link:

[Link to publication record in Edinburgh Research Explorer](#)

### Document Version:

Peer reviewed version

### Published In:

Bayesian analysis

### General rights

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

### Take down policy

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact [openaccess@ed.ac.uk](mailto:openaccess@ed.ac.uk) providing details, and we will remove access to the work immediately and investigate your claim.



# Colombian Women's Life Patterns: A Multivariate Density Regression Approach

Sara Wade<sup>\*</sup>, Raffaella Piccarreta<sup>†</sup>, Andrea Cremaschi<sup>‡</sup>, and Isadora Antoniano-Villalobos<sup>§</sup>

**Abstract.** Women in Colombia face difficulties related to the patriarchal traits of their societies and well-known conflict afflicting the country since 1948. In this critical context, our aim is to study the relationship between baseline socio-demographic factors and variables associated to fertility, partnership patterns, and work activity. To best exploit the explanatory structure, we propose a Bayesian multivariate density regression model, which can accommodate mixed responses with censored, constrained, and binary traits. The flexible nature of the models allows for nonlinear regression functions and non-standard features in the errors, such as asymmetry or multi-modality. The model has interpretable covariate-dependent weights constructed through normalization, allowing for combinations of categorical and continuous covariates. Computational difficulties for inference are overcome through an adaptive truncation algorithm combining adaptive Metropolis-Hastings and sequential Monte Carlo to create a sequence of automatically truncated posterior mixtures. For our study on Colombian women's life patterns, a variety of quantities are visualised and described, and in particular, our findings highlight the detrimental impact of family violence on women's choices and behaviors.

**MSC2020 subject classifications:** Primary 62G07, 62G08; secondary 62N01, 62P25.

**Keywords:** Bayesian nonparametrics, adaptive truncation, sequential Monte Carlo, time-to-event, non-informative censoring.

## 1 Introduction

Colombian women face difficulties that are quite typical in Latin American countries, particularly related to the patriarchal traits of their society. Nonetheless, the welfare of Colombian women is possibly more critical due to the conflict between state military forces, paramilitaries, and guerrilla groups that has afflicted the country since 1948. As underlined by [Gimenez Duarte et al.](#), dramatic subnational inequalities exist in every indicator, especially within low-income, low-education, and rural populations. Reinforcing constraints, such as “limited and gender-unequal economic opportunities, exclusion from quality endowments among marginalized populations, and social norms and gender roles that relegate unpaid care work to women and tolerate violence against them

---

<sup>\*</sup>School of Mathematics, University of Edinburgh, Edinburgh, UK [sara.wade@ed.ac.uk](mailto:sara.wade@ed.ac.uk)

<sup>†</sup>Department of Decision Sciences, Dondena Research Centre, Bocconi Institute for Data Science and Analytics (BIDSA), Bocconi University, Milan, Italy [raffaella.piccarreta@unibocconi.it](mailto:raffaella.piccarreta@unibocconi.it)

<sup>‡</sup>Yale-NUS College, Singapore [andrea.cremaschi@yale-nus.edu.sg](mailto:andrea.cremaschi@yale-nus.edu.sg)

<sup>§</sup>Department of Environmental Sciences, Informatics and Statistics, Ca' Foscari University of Venice and BIDSA, Italy [isadora.antoniano@unive.it](mailto:isadora.antoniano@unive.it)

(emotional, physical and sexual), affect young women’s choices and actions with respect to life plans and fertility decisions” (Gimenez Duarte et al., 2015, p. 5). In particular, despite significant progress since 2000, teenage pregnancy rates in Colombia are still very high. The majority of teenage pregnancies remain unplanned, signaling a lack of opportunity and agency for young girls. Different studies discuss the detrimental effects of teenage pregnancy (see e.g., Gimenez Duarte et al., 2015; Azevedo et al., 2012) and its socio-demographic drivers, such as poverty, low levels of education, and living in rural areas. In such a critical context, we are interested in studying women’s life events, focusing on the interplay between sexual initiation (debut), fertility, partnership, and participation in the labor market. Thus, rather than focusing on a specific life event, as in previous relevant studies (e.g. Gimenez Duarte et al., 2015; Azevedo et al., 2012; Restrepo Martínez et al., 2017), we adopt a broader perspective, considering a collection of events describing transition to adulthood and their relation with a set of structural baseline characteristics of the women’s environment and family. Besides some of the well known relevant factors, such as cohort, region, and area (urban or rural) of residence, we also study whether a violent family context contributes to shape transition to adulthood and possibly impairs women’s agency.

To this purpose, we analyze data arising from the survey conducted in Colombia in 2010 as a part of the Demographic and Health Survey Program (DHS, <https://www.dhsprogram.com>). The data are cross-sectional, thus, no follow-up information on the life events of interest is recorded. Specifically, information is available on the age when the considered focal events (sexual debut, marriage or cohabitation, motherhood) were experienced for the first time, whereas work information concerns only the employment status of the woman (working or not) at the moment of the interview. Thus, we jointly analyze response variables with different levels of measurements (times at event and binary variables). Additionally, the focal events may not have been experienced (right-censoring) and are subject to constraints, e.g. motherhood can only occur after sexual debut. Furthermore, the available set of baseline explanatory variables is limited, so that heterogeneity may be present which would not be properly captured by a parametric model. This encourages the use of a flexible model to best exploit the explanatory structure without imposing possibly penalizing constraints. Additionally, such a model can encompass competing sociological theories which may be relevant in different subpopulations.

The data present various features that challenge existing parametric and semiparametric models (e.g. Korsgaard et al., 2003; Jara et al., 2010; Hanson and Johnson, 2004; Kottas and Gelfand, 2001). First, some women postpone the events to relatively late in life, which induces right-skewed distributions. Also, the joint relationships between the age-at-event variables show different patterns, with gaps of various lengths between events. Moreover, these behaviors change depending on the covariates. Modeling such dependence structure is an ambitious task, requiring a model that allows for i) non-linear response curves, ii) non-normal distributions whose features may change with the covariates, iii) multivariate response and covariates of mixed nature, and iv) censoring and constraints of the responses. To the best of our knowledge, a model that can simultaneously deal with these issues does not exist.

We propose a Bayesian multivariate density regression model that extends the univariate model of [Antoniano-Villalobos et al. \(2014\)](#) to the case of multiple mixed-type responses with censoring and constraints. This approach is promising for our data, due to its ability to capture their peculiar features. Our infinite mixture model has interpretable covariate-dependent weights constructed through normalization, allowing for combinations of categorical and numerical covariates. In addition, the multivariate approach permits us to study the joint relationship between the response variables, for example by considering one response conditioned on the others. With data on over 10,000 women and a multivariate response and covariate, the Markov chain Monte Carlo (MCMC) algorithm originally proposed for the univariate model becomes unsuitable. We therefore propose an algorithm for posterior inference that extends the adaptive truncation scheme of [Griffin \(2016\)](#).

The paper is structured as follows. Section 2 describes the data. The model and posterior simulation algorithm are presented in Sections 3 and 4, respectively. The results for the data on Colombian women are analyzed in Section 5. Section 6 summarizes and concludes. In addition, the Supplementary Material (SM) includes derivations and details for predictive inference, as well as additional results for both the simulated data example and the case study.

## 2 The data

The DHS Program collects and disseminates data on random samples of households selected from random clusters from a national sampling frame<sup>1</sup>. The 2010 survey in Colombia was conducted by the Profamilia association, and we refer to the final report for a detailed description of its features ([Ojeda et al., 2011](#)). Since all women of childbearing potential (i.e. aged 13-49) in the same household were interviewed, we randomly select at most one case from each household to avoid unwanted dependencies.

To describe the characteristics of fertility and partnership patterns, we consider the discrete variables recording the ages at *Sexual Debut* ( $Z_1$ ), at *Union* (first marriage or cohabitation,  $Z_2$ ), and at *First Child* ( $Z_3$ ). *Work Status* is recorded as a binary variable ( $Z_4$ ) indicating whether the respondent worked in the 12 months before the interview. We exclude women who gave inconsistent information, namely, those who report the birth of the first child as preceding the first sexual intercourse, and those who report union with a partner but for whom sexual intercourse never occurred. We also filter out women who experienced sexual violence or were forced to have sex in exchange for money, since their union and childbearing choices may be related to the experienced violence. By the same reasoning, we remove women who were forced to use contraceptive methods. Thus, we attempt to focus as much as possible on life choices and plans rather than on events imposed by circumstances, even if the latter may be unknown and unmeasured, so that the observed events may not necessarily reflect choices.

---

<sup>1</sup>Note that the data arise from a complex survey design, and thus have associated weights with a complicated structure. Moreover, additional post-stratification is carried out to adjust the weights for various factors, such as the total number of women interviewed in each municipality, non-response, etc. (see [Ojeda et al., 2011](#), for full details on the survey design and weighting scheme for the Colombian survey). A discussion on the suitability of accounting for such sampling weights is offered in the SM.

Age	Sexual Debut		Union		First Child	
	Censored	Observed	Censored	Observed	Censored	Observed
15–19	1,144	1,053	1,818	379	1,837	360
20–29	238	3,475	1,358	2,355	1,323	2,390
30–39	51	2,597	378	2,270	326	2,322
40–49	55	2,127	281	1,901	216	1,966
	1,488	9,252	3,835	6,905	3,702	7,038

Table 1: Cross-tabulation of age groups and censored data.

We focus on the relationship between the responses and some baseline socio-demographic factors. First, we consider the woman’s *Age* (in years) at the moment of interview. We focus on women aged 15 or more, as most younger women had not yet experienced any event at the time of the survey. Next, we include the type of *Area* (urban or rural) where the respondent lives, as well as her *Region*. Following the partition used in the DHS dataset (as well as in the report by [Ojeda et al., 2011](#)) we consider the five regions Atlantica, Oriental, Central, Pacifica, and Territorios Nacionales<sup>2</sup>; the capital Bogota, located in the Oriental region, is treated as a sixth region because of its peculiar features in terms of social and economic development. A map of Colombia and the considered regions is reported in the SM. Since information is only available on the current region of residence and on the age when respondents moved there, we limit attention to women who were raised in the current region at least from the age of 6, to properly account for regional effects. Moreover, to assess the respondent’s well-being in her original family, we consider whether in her childhood she was disciplined using *Physical Punishment* (spanking, hitting, pushing, throwing water), and whether she was exposed to *Parental Domestic Violence* and ever witnessed her father beating her mother. All respondents missing at least one explanatory or response variable are excluded from the dataset.

Even if the DHS dataset is very rich, including other covariates is not straightforward. Most of the variables refer to the moment of interview, and thus cannot be considered as antecedents of the focal events. For example, although it would be interesting to include information regarding education and wealth, only the highest level of education attained and the wellness of the respondent’s family at the moment of interview are available. Another relevant aspect that could be taken into account concerns women’s ethnicity. However, most (about 80%) of the interviewed women do not recognize themselves as part of an ethnic minority. Furthermore, those who do belong to a heterogeneous variety of ethnic groups, none of which are sufficiently represented in the sample. We therefore exclude ethnic minorities from our study.

Our final dataset consists of  $n = 10740$  women. Table 1 reports a summary of the number of censored cases for the first three response variables within age groups. Figures 1 and 2 offer some insights about the distinctive features of the densities of the ages at events and of their relationships (smoothed regressions) with the *Age* at interview across different covariates values. Even if the displayed results only relate to

<sup>2</sup>The former region “Territorios Nacionales” includes the two more easterly regions, Orinoquia and Amazonia, created in 1991.

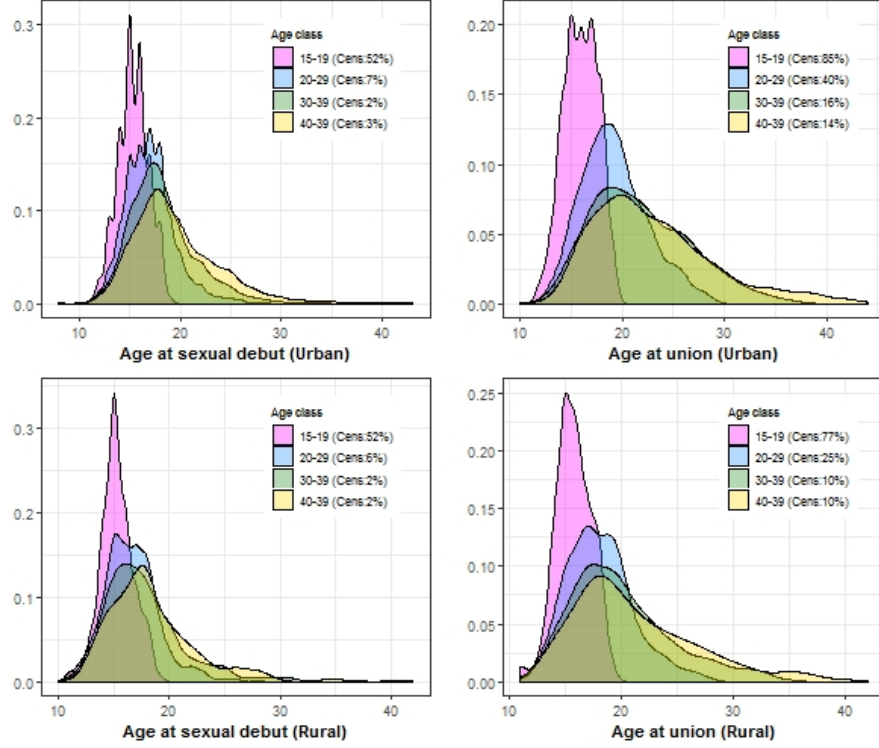


Figure 1: Kernel density estimates of (non-censored) ages at events conditioned to *Age* (at interview, in groups) and area of residence (Urban or Rural).

non-censored cases, they emphasize the difficulties implied by modeling the dependence structures. To achieve such an ambitious task, in the next section we propose a model that allows for non-linear response curves (see e.g. Figure 2), non-normal distributions whose features may change with the covariates (see e.g. Figure 1), multivariate response and covariates of mixed nature, and censoring of the responses<sup>3</sup>.

### 3 Bayesian nonparametric density regression

We develop a Bayesian nonparametric mixture model that can capture the relationship between  $n$  conditionally independent  $d$ -dimensional response vectors,  $\mathbf{Z}_i$ , and a vector  $\mathbf{x}_i^*$  of predictors. To simplify notation, whenever possible we drop the sub-index  $i$  indicating individual observations. The predictors  $\mathbf{x}^* = (x_1, \dots, x_p, x_{p+1}^*, \dots, x_{q^*}^*)$  may be of mixed nature. Without loss of generality, we assume that the first  $p$  are numerical while the

<sup>3</sup>Please note that we refer specifically to *random* (or non-informative) censoring, i.e. to the case when each subject has a censoring time that is statistically independent of the age at a given response event.

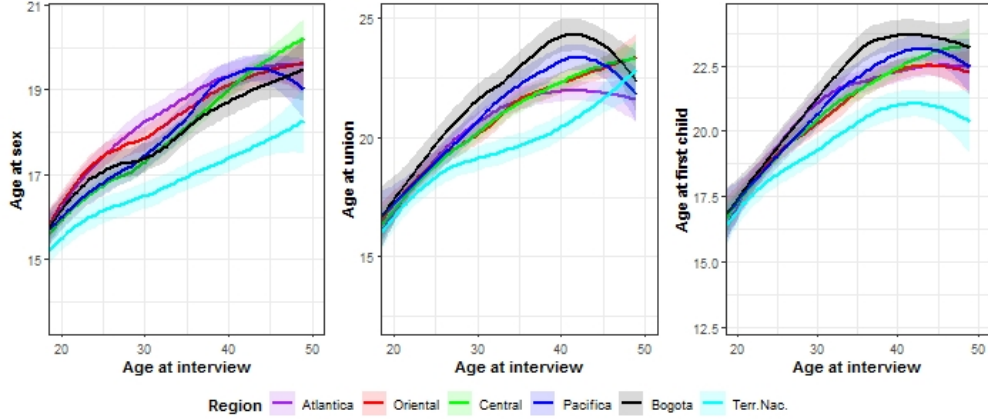


Figure 2: Relation (smoothed regression) between (non-censored) ages at events and *Age*.

rest are categorical. As is common in regression models, we expand the categorical predictors with binary dummy variables and let  $\mathbf{x} = (x_1, \dots, x_p, x_{p+1}, \dots, x_q)$ , where  $q = p + \sum_{k=p+1}^{q^*} (R_k - 1)$  and  $R_k$  denotes the number of categories of  $x_k^*$ . The observed responses are also of mixed nature. For example, in our application, we consider two types of responses: three positive integer-valued variables with possible censoring and constraints, representing the ages at events, and one binary variable indicating work status. In this case, we refer to the density of the mixed response  $\mathbf{Z} = (Z_1, \dots, Z_d)$  with respect to the appropriate measure, e.g. Lebesgue or counting measure, for each response type.

To frame our model within existing literature, we review some related contributions. Bayesian nonparametric mixture models (Lo, 1984) are useful tools for density estimation, due to their attractive balance between flexibility and smoothness and ability to recover a wide range of densities (Ghosh and Ramamoorthi, 2003, Chapter 5). Further developments for conditional density estimation, also known as density regression, can be found in the pioneering works of Müller et al. (1996) and MacEachern (1999). Extensions of the former for categorical responses can be found in Shahbaba and Neal (2009). We focus on the latter work of MacEachern (1999), which extends the Bayesian nonparametric mixture model by allowing the mixing measure to depend on the covariates. This yields flexible density regression, where the entire density and not only the mean is regressed on the covariates. Several approaches exist in literature to specify the covariate-dependent mixing measure, but it is not clear how to choose between them. Examples include single- $p$  dependent Dirichlet processes (MacEachern, 2000; De Iorio et al., 2004), with covariate-dependent component parameters but single weights. Alternatively, numerous proposals have been introduced for covariate-dependent weights (Griffin and Steel, 2006; Dunson and Park, 2008; Rodriguez and Dunson, 2011, to name a few). In terms of the support and consistency properties of the model, the distinc-



tion is irrelevant (see e.g. [Barrientos et al., 2012, 2017](#); [Pati et al., 2013](#)). However, the analysis of [Wade et al. \(2014b\)](#) suggests that prediction issues may arise at moderate sample sizes for single weights mixtures with linear predictors, when the true relation between response and covariates is non-linear. Given the complex nature of the application that we are considering and the lack of theory justifying transformations that would simplify the shape of the relation between the variables, we have decided to consider a linear predictor approach. Therefore, in this work, we build on the construction of the covariate-dependent weights developed by [Antoniano-Villalobos et al. \(2014\)](#), which allows for combinations of continuous and discrete covariates and favors interpretability.

We require extending the model to multivariate responses of mixed type with possible censoring and constraints. An appealing approach for this relies on a latent Gaussian representation, which provides a simple construction for dependence of the multivariate mixed-type data through the full covariance matrix of the latent Gaussian variables. Moreover, Bayesian inference can be carried out through Gibbs sampling and data augmentation techniques. A Bayesian parametric model based on this idea was proposed by [Korsgaard et al. \(2003\)](#) for multivariate data combining Gaussian, right-censored Gaussian, ordinal, and binary traits. To increase model flexibility, Bayesian nonparametric versions were proposed by [De Yoreo et al. \(2017\)](#) for mixed ordinal and nominal data, by [De Yoreo and Kottas \(2018a\)](#) for multivariate ordinal regression (see [De Yoreo and Kottas, 2018b](#), for a dynamic extension) and by [Papageorgiou et al. \(2015\)](#) for mixed-type spatial data in an epidemiological context. Due to the increased flexibility of nonparametric mixtures, the cut-offs used to define the discrete data from the latent Gaussian variables can be fixed and not estimated or inferred. Moreover, [Canale and Dunson \(2011\)](#) show that Bayesian nonparametric mixtures for discrete data (specifically counts) based on latent Gaussian variables can approximate and consistently estimate a wider range of distributions than mixtures based on discrete distributions, e.g. Poisson or multinomial (see also [Kottas et al. \(2005\)](#); [Bao and Hanson \(2015\)](#) in the context of multivariate ordinal data and [Jara et al. \(2007\)](#) for binary response regression). Another relevant extension is the Bayesian semiparametric model of [Jara et al. \(2010\)](#) for multivariate doubly-censored data indicating time to event, based on a log transformation linking the observed responses to the latent Gaussian variables. When modeling time-to-event data, the log transformation is more appropriate than others, notably truncation. This allows recovery of the underlying structure with fewer and more interpretable components with possibly heavy right tails. Recently, [Norets and Pelenis \(2020\)](#) demonstrated that optimal adaptive estimation of mixed discrete-continuous distributions can be achieved via the latent Gaussian mixture approach.

We combine some of these ideas to build a model which can deal with the challenges presented by the data. We adopt the latent Gaussian approach, associating to each response variable  $Z_\ell$  a latent real-valued  $Y_\ell$ . Specifically, an observed value  $z_\ell$  of the response  $Z_\ell$  is linked to the realization  $\mathbf{y} = (y_1, \dots, y_d)$  of the latent  $\mathbf{Y} = (Y_1, \dots, Y_d)$ , through a function  $h_\ell$  whose characteristics depend on the nature of the observable. Examples of transformations for different response types include:

$$\begin{aligned} z_\ell &= h_\ell(\mathbf{y}, \mathbf{x}) = y_\ell, & \text{for } z_\ell \in \mathbb{R}, \\ z_\ell &= h_\ell(\mathbf{y}, \mathbf{x}) = \lfloor \exp(y_\ell) \rfloor, & \text{for } z_\ell \in \mathbb{N}, \end{aligned}$$



$$z_\ell = h_\ell(\mathbf{y}, \mathbf{x}) = \sum_{a=1}^{A_\ell-1} \mathbb{1}_{[\alpha_{\ell,a}, \infty)}(y_\ell), \quad \text{for } z_\ell \in \{0, 1, 2, \dots, A_\ell - 1\},$$

where  $\lfloor \cdot \rfloor$  denotes the floor function, and  $\mathbb{1}_B(y)$  denotes the indicator function taking the value one when  $y \in B$ . Note that the last case considers an ordinal response with  $A_\ell$  categories and fixed cutoffs of  $\alpha_{\ell,1} < \dots < \alpha_{\ell,A_\ell-1}$ . In these examples, the functions  $h_\ell$  do not depend on  $\mathbf{x}$  or  $y_{\ell'}$  for  $\ell' \neq \ell$ , but they may, for example when accounting for censored or constrained responses, as for the case study described in Section 5.

The basic building block for our model is the multivariate multiple linear regression:

$$\mathbf{Y}|\mathbf{x}, \boldsymbol{\beta}, \boldsymbol{\Sigma} \stackrel{ind}{\sim} N_d(\mathbf{y}|\mathbf{x}\boldsymbol{\beta}, \boldsymbol{\Sigma}),$$

where  $\boldsymbol{\beta}$  is a  $(q+1) \times d$  matrix of regression parameters and  $\boldsymbol{\Sigma}$  is a  $d \times d$  covariance matrix. Slightly abusing notation,  $\mathbf{x} = (1, x_1, \dots, x_q)$  denotes the vector of observed covariate values extended by a unitary entry. As previously discussed, this parametric model is not flexible enough to capture the complex dependence structures contained in the data. We therefore extend the nonparametric density regression framework introduced by [Antoniano-Villalobos et al. \(2014\)](#) to model the  $\mathbb{R}^d$ -valued latent variable  $\mathbf{Y}$ :

$$f_{\mathbf{P}_\mathbf{x}}(\mathbf{y}|\mathbf{x}) = \sum_{j=1}^{\infty} w_j(\mathbf{x}) N_d(\mathbf{y}|\mathbf{x}\boldsymbol{\beta}_j, \boldsymbol{\Sigma}_j), \quad \text{with} \quad w_j(\mathbf{x}) = \frac{w_j g(\mathbf{x}|\boldsymbol{\psi}_j)}{\sum_{j'=1}^{\infty} w_{j'} g(\mathbf{x}|\boldsymbol{\psi}_{j'})}. \quad (1)$$

This model results from considering a mixture

$$f_{\mathbf{P}_\mathbf{x}}(\mathbf{y}|\mathbf{x}) = \int N_d(\mathbf{y}|\mathbf{x}\boldsymbol{\beta}, \boldsymbol{\Sigma}) d\mathbf{P}_\mathbf{x}(\boldsymbol{\theta}),$$

where  $\boldsymbol{\theta} = (\boldsymbol{\beta}, \boldsymbol{\Sigma})$  and a nonparametric prior is assigned to the set of covariate-dependent mixing measures  $\mathbf{P}_\mathbf{x}$ , which places mass one on the set of discrete probability measures:

$$\mathbf{P}_\mathbf{x} = \sum_{j=1}^{\infty} w_j(\mathbf{x}) \delta_{\boldsymbol{\theta}_j}.$$

Here,  $\delta_\theta$  denotes the Dirac-delta function with unit mass at  $\theta$ . For computational purposes and to ensure convergence of the normalizing constant in  $w_j(\mathbf{x})$ , it is convenient to adopt a stick-breaking representation for the weights, setting  $w_1 = v_1$  and  $w_j = v_j \prod_{j' < j} (1 - v_{j'})$ , for  $j > 1$ , where  $v_j \stackrel{ind}{\sim} \text{Beta}(\zeta_{j,1}, \zeta_{j,2})$ . The parameters of the local linear regression components,  $\boldsymbol{\theta}_j$ , and of the covariate-dependent weights,  $\boldsymbol{\psi}_j$ , are assumed to be independent and identically distributed according to a base measure  $\mathbf{P}_0$  and independent of the weights. Together with the functions  $h_\ell$  linking the latent variables with the responses, this defines the likelihood structure for the observed data.

In this model, the regression parameters  $\boldsymbol{\beta}_j$  and  $\boldsymbol{\Sigma}_j$  capture the local linear relation between the latent response and covariates, with normal errors; whereas the  $\boldsymbol{\psi}_j$  determine, through  $g$ , how the influence of each local component to the overall model changes

across the covariate space. This deals with situations when the stochastic relation between  $\mathbf{y}$  and  $\mathbf{x}$  is too complicated to be captured by a single parametric model. It can also be used when the population is assumed to be constituted by an unknown number of (covariate-dependent) groups such that, within each group, a linear regression model provides a good description of the data. While identifiability issues may prevent the individuation of such groups, this intuition can help in understanding the elements of the model.

Note that the Bayesian nonparametric model for the joint density of  $\mathbf{y}$  and  $\mathbf{x}$  introduced by Müller et al. (1996) for density regression, taking the form

$$f_{\mathbf{P}}(\mathbf{y}, \mathbf{x}) = \sum_{j=1}^{\infty} w_j g(\mathbf{x}|\boldsymbol{\psi}_j) N_d(\mathbf{y}|\mathbf{x}\boldsymbol{\beta}, \boldsymbol{\Sigma}), \text{ with } \mathbf{P} = \sum_{j=1}^{\infty} w_j \delta_{(\boldsymbol{\theta}_j, \boldsymbol{\psi}_j)}, \quad (2)$$

results in a conditional density coinciding with equation (1). However, an important difference is that in the joint mixture model, posterior inference for the parameters  $(w_j, \boldsymbol{\theta}_j, \boldsymbol{\psi}_j)$  is based on the joint likelihood in (2); whereas, for our model, it is based directly on the conditional likelihood of interest. Furthermore, we emphasize that the converse is not true: our conditional density model in (1) does not imply the joint density model in (2). This can be easily seen by constructing a joint density model as the product of (1) and any, say parametric, marginal density model for  $\mathbf{x}$ . This is a valid construction, which nonetheless recovers the joint model in (2) only when the marginal has the form:

$$f_{\mathbf{P}}(\mathbf{x}) = \sum_{j=1}^{\infty} w_j g(\mathbf{x}|\boldsymbol{\psi}_j).$$

This is an important concept, as it highlights that the form chosen for  $g$  does not imply a modeling of the distribution for covariates, which may indeed be fixed. The choice and shape of this kernel, however, defines how the conditional distribution changes as  $\mathbf{x}$  varies (given the parameters  $\boldsymbol{\psi}$ ). Thus, it determines the amount of information borrowed when making inference at unobserved points in the space of covariates. By choosing model (1) we maintain the same natural and interpretable structure for the weights of the joint mixture model, but exploit all the information available in the data to learn about the relation between  $\mathbf{x}$  and  $\mathbf{y}$ , thus improving the quality of the estimation for the conditional distribution (see Wade et al., 2014a), which is the main focus of our application. Clearly, a practitioner interested also in capturing the structure of the random covariates would require a different approach.

The covariate-dependent weight  $w_j(\mathbf{x})$  represents the probability that an observation with a covariate value  $\mathbf{x}$  is allocated to the  $j$ -th regression component. Such probability can be decomposed into the unconditional probability  $w_j$  that the parametric model  $j$  fits an individual observation, and the likelihood  $g(\mathbf{x}|\boldsymbol{\psi}_j)$  that an individual allocated to the  $j$ -th component is characterized by a covariate value  $\mathbf{x}$ . The  $g(\cdot|\boldsymbol{\psi})$  can be defined to accommodate different types of covariates. We adopt a factorizable structure:

$$g(\mathbf{x}|\boldsymbol{\psi}) = \prod_{k=1}^q g(x_k|\psi_k), \quad \text{where} \quad g(x_k|\psi_k) = \begin{cases} N(x_k|\mu_k, \tau_k^{-1}) & \text{for } k = 1, \dots, p, \\ \text{Bern}(x_k|\rho_k) & \text{for } k = p+1, \dots, q, \end{cases}$$

with  $\psi_k = (\mu_k, \tau_k)$  for  $k = 1, \dots, p$ , and  $\psi_k = \rho_k$  for  $k = p + 1, \dots, q$ . The use of distribution kernels guarantees convergence, for all  $\mathbf{x}$ , of the denominator in equation (1). For the unconditional probability  $w_j$ , different choices of the stick-breaking parameters  $(\zeta_{j,1}, \zeta_{j,2})$  result in different nonparametric priors (see Ishwaran and James, 2001). For instance, if  $(\zeta_{j,1}, \zeta_{j,2}) = (1, \zeta)$ , the prior on the weights  $w_j$  corresponds to that obtained from a Dirichlet process prior. The base measure is chosen as  $\mathbf{P}_0(\boldsymbol{\beta}, \boldsymbol{\Sigma}, \boldsymbol{\psi}) = \mathbf{P}_0(\boldsymbol{\beta}|\boldsymbol{\Sigma})\mathbf{P}_0(\boldsymbol{\Sigma})\mathbf{P}_0(\boldsymbol{\mu}|\boldsymbol{\tau})\mathbf{P}_0(\boldsymbol{\tau})\mathbf{P}_0(\boldsymbol{\rho})$ . We use the conjugate matrix-variate Normal-Inverse Wishart for the regression parameters:  $\mathbf{P}_0(\boldsymbol{\beta}|\boldsymbol{\Sigma}) = \text{MN}_{(q+1) \times d}(\boldsymbol{\beta}_0, \mathbf{U}, \boldsymbol{\Sigma})$ , where  $\boldsymbol{\beta}_0$  is a  $(q+1) \times d$  matrix and  $\mathbf{U}$  is a  $(q+1) \times (q+1)$  positive definite matrix; and  $\mathbf{P}_0(\boldsymbol{\Sigma}) = \text{IW}(\boldsymbol{\Sigma}_0, \nu)$ , where  $\boldsymbol{\Sigma}_0$  is a  $d \times d$  positive definite matrix and  $\nu > 0$ . Notice that the Inverse Wishart assigns prior mass to full covariance matrices. Other prior specifications can be used to allow for other types of covariance structures, e.g. a product of Inverse Gammas for diagonal covariance matrices or a G-Wishart for sparse precision matrices. As for  $\boldsymbol{\beta}$ , we are assuming a structured dependence, allowing for efficient computations through Kronecker products and a reduced number of hyperparameters compared to a full Gaussian distribution. Alternatively, a multivariate Gaussian distribution could be used, assuming independence between columns. To complete the specification of the base measure, we set:  $\mathbf{P}_0(\boldsymbol{\mu}|\boldsymbol{\tau}) = \prod_{k=1}^p \text{N}(\mu_k|\mu_{0,k}, (u_k \cdot \tau_k)^{-1})$ ,  $\mathbf{P}_0(\boldsymbol{\tau}) = \prod_{k=1}^p \text{Gamma}(\tau_k|\alpha_k, \gamma_k)$ , and  $\mathbf{P}_0(\boldsymbol{\rho}) = \prod_{k=p+1}^q \text{Beta}(\rho_k|\boldsymbol{\varrho}_k)$ , where  $\boldsymbol{\varrho}_k = (\varrho_{k,1}, \varrho_{k,2})$ .

In the next section, we describe an adaptive truncation algorithm allowing posterior inference for our model. The algorithm is general and only requires specific adjustments depending on the  $h_\ell$  functions linking the observed responses with their latent counterparts.

## 4 Adaptive truncation algorithm

To scale appropriately with the sample size and data dimensions, we devise an algorithm for posterior inference based on a finite truncation of the mixture, where the number of components is allowed to increase adaptively to obtain a good approximation of the infinite-dimensional posterior. The truncated latent model with  $J$  components is:

$$f_{\mathbf{P}_x^J}(\mathbf{y}|\mathbf{x}) = \sum_{j=1}^J w_j^J(\mathbf{x}) \text{N}_d(\mathbf{y}|\mathbf{x}\boldsymbol{\beta}_j, \boldsymbol{\Sigma}_j), \quad (3)$$

where the weights follow the re-normalized stick breaking construction:

$$w_j^J(\mathbf{x}) = \frac{w_j g(\mathbf{x}|\boldsymbol{\psi}_j) / \sum_{j=1}^J w_j}{\sum_{j'=1}^J w_{j'} g(\mathbf{x}|\boldsymbol{\psi}_{j'}) / \sum_{j=1}^J w_j} = \frac{w_j g(\mathbf{x}|\boldsymbol{\psi}_j)}{\sum_{j'=1}^J w_{j'} g(\mathbf{x}|\boldsymbol{\psi}_{j'})}. \quad (4)$$

Notice that the normalizing constant  $\sum_{j=1}^J w_j$  in (4) cancels out. To ease notation, we use  $w_j(\mathbf{x})$  to denote the truncated covariate-dependent weights, dropping the superscript  $J$  when the truncation level is clear. Due to the exponential decay of the weights, for large enough  $J$ , the truncated model (3) provides a close approximation to the infinite mixture model. Alternative truncation methods could be considered, notably the

popular truncated stick breaking method (Ishwaran and James, 2001) where  $v_J = 1$ . However, re-normalized stick-breaking may provide a better finite-dimensional approximation by evenly distributing the remaining mass across components, as opposed to assigning all remaining mass to the last component in truncated stick-breaking.

The proposed algorithm is based on the adaptive truncation scheme developed by Griffin (2016), extended for density regression and mixed type responses. It consists of two main steps, namely an MCMC step for a fixed truncation level,  $J_0$ , followed by a sequential Monte Carlo (SMC) step used to increase the number of components of the mixture. The first step produces  $M$  posterior draws  $(\mathbf{w}_{1:J_0}^m, \boldsymbol{\theta}_{1:J_0}^m, \boldsymbol{\psi}_{1:J_0}^m, \mathbf{y}_{1:n}^m)_{m=1}^M$ , which are then used as particles in the SMC step. We provide a concise summary below, with software and full details provided through the authors' GitHub repository<sup>4</sup> and accompanying documentation.

**MCMC for fixed truncation.** Since the truncation level  $J_0$  is fixed throughout this step, we omit it from the notation, writing  $\mathbf{w} = w_{1:J_0}$ ,  $\boldsymbol{\theta} = \boldsymbol{\theta}_{1:J_0}$ , and  $\boldsymbol{\psi} = \boldsymbol{\psi}_{1:J_0}$ . Similarly, the observed response is denoted by  $\mathbf{z} = (\mathbf{z}_1, \dots, \mathbf{z}_n)$ , with  $\mathbf{z}_i = (z_{i,1}, \dots, z_{i,d})$ , and analogously for the covariates  $\mathbf{x}$  and the latent  $\mathbf{y}$ . The approximate posterior given the sample  $(\mathbf{x}, \mathbf{z})$  of size  $n$ , using the truncated likelihood (3), takes the form:

$$\mathbf{P}_{J_0}^n(\mathbf{w}, \boldsymbol{\psi}, \boldsymbol{\theta}, \mathbf{y} | \mathbf{z}, \mathbf{x}) \propto \mathbf{P}_{J_0}(\mathbf{w}, \boldsymbol{\psi}, \boldsymbol{\theta}) \prod_{i=1}^n \sum_{j=1}^{J_0} w_j(\mathbf{x}_i | \boldsymbol{\psi}_j) \mathcal{N}_d(\mathbf{y}_i | \mathbf{x}_i \boldsymbol{\beta}_j, \boldsymbol{\Sigma}_j) \prod_{\ell=1}^d \mathbb{1}_{\{z_{i,\ell}\}}(h_{i,\ell}),$$

where  $\mathbf{P}_{J_0}(\mathbf{w}, \boldsymbol{\psi}, \boldsymbol{\theta})$  indicates the restriction of the prior (as detailed in Section 3) to the parameters in the truncated space. Dependence  $w_j(\mathbf{x}) = w_j(\mathbf{x} | \boldsymbol{\psi}_j)$  of the weights on the parameters has been made explicit. Moreover, the functions  $h_{i,\ell} = h_\ell(\mathbf{y}_i, \mathbf{x}_i)$  linking the latent variables to the observed responses are tailored to the specific application in Section 5.

Due to lack of conjugacy, we resort to a generic Metropolis-within-Gibbs scheme to perform posterior sampling, that updates blocks of parameters adaptively. The adaptive random walk algorithm used here, based on Algorithm 6 in Griffin and Stephens (2013), adapts the proposal covariance matrix to achieve both a specified average acceptance rate ( $a_0 = 0.234$ ) and a proposal covariance matrix equal to  $2.4^2/\mathbf{p}$  times the posterior covariance matrix,  $\mathbf{p}$  being the dimension of the parameter block of interest. These criteria have been shown to be optimal in many settings (Roberts et al., 1997; Roberts and Rosenthal, 2001). In more detail, suppose that we want to sample a block of parameters  $\boldsymbol{\phi}$  of dimension  $\mathbf{p}$  from a distribution with probability density function  $Q$ . In order to utilise the adaptive random walk algorithm, we first consider a transformation  $t(\boldsymbol{\phi})$  that has full support on  $\mathbb{R}^{\mathbf{p}}$ . At each iteration  $m$ , we propose a new  $\boldsymbol{\phi}^*$  such that:

$$\mathbf{t}^* \equiv t(\boldsymbol{\phi}^*) = t(\boldsymbol{\phi}^{m-1}) + \boldsymbol{\epsilon}, \text{ with } \boldsymbol{\epsilon} \sim \mathcal{N}(0, \boldsymbol{\xi}^{m-1}), \quad (5)$$

where  $\boldsymbol{\xi}^{m-1}$  is the adaptive covariance matrix. We accept  $\boldsymbol{\phi}^m = \boldsymbol{\phi}^*$  with probability equal to the minimum between 1 and the ratio:

$$a(\boldsymbol{\phi}^*, \boldsymbol{\phi}^{m-1}) = \frac{Q(\boldsymbol{\phi}^*)}{Q(\boldsymbol{\phi}^{m-1})} \frac{|\mathcal{J}_t(\boldsymbol{\phi}^{m-1})|}{|\mathcal{J}_t(\boldsymbol{\phi}^*)|}, \quad (6)$$

<sup>4</sup>[https://github.com/sarawade/BNPDensityRegression\\_AdaptiveTruncation](https://github.com/sarawade/BNPDensityRegression_AdaptiveTruncation)

with  $|\mathcal{J}_t(\phi)|$  denoting the determinant of the Jacobian of the transformation.

Transformations of  $\beta_j$ ,  $\mu_j$ ,  $\tau_j$ ,  $\rho_j$ , and  $\mathbf{v}_j$  are straightforward through identity, log, and logit functions. Instead, transformations of  $\Sigma_j$  and  $\mathbf{y}_i$  are more involved. For each  $\Sigma_j$ , we consider a vectorization of a decomposition of the matrix,  $\Sigma_j = \mathbf{L}_j \mathbf{D}_j \mathbf{L}_j^\top$ , where  $\mathbf{L}_j$  is a lower triangular matrix with unit entries on the diagonal and  $\mathbf{D}_j$  is a diagonal matrix with positive entries, and we take the log of the diagonal entries. In this case, the proposed  $\Sigma_j^*$  can be obtained from the proposed  $\mathbf{t}^*$  in equation (5) by inverting this transformation. In addition, the determinant of the Jacobian, which is required in the acceptance ratio in (6), depends only on the diagonal elements  $D_{j,\ell,\ell}$  of the matrix  $\mathbf{D}_j$ , specifically,  $|\mathcal{J}_t(\Sigma_j)| = \prod_{\ell=1}^d 1/D_{j,\ell,\ell}^{d+1-\ell}$ . For each latent vector  $\mathbf{y}_i = (y_{i,1}, \dots, y_{i,d})$ , the terms  $h_\ell(\mathbf{y}_i, \mathbf{x}_i) = z_{i,\ell}$  define constrained regions for the latent  $\mathbf{y}_i$ , such that  $y_{i,\ell} \in (l_{i,\ell}, u_{i,\ell})$ , which are provided for the case study in Section 5. We assume that an appropriate ordering of the responses leads to bounds  $(l_{i,\ell}, u_{i,\ell})$  that may in general depend on  $y_{i,\ell'}$  for  $\ell' < \ell$ . This allows us to define a sequential logistic transformation  $t(y_{i,\ell}; \mathbf{y}_{i,1:\ell-1})$  for  $\ell = 1, \dots, d$ , based on the bounds  $(l_{i,\ell}, u_{i,\ell})$ . From the proposed  $\mathbf{t}^*$  in equation (5), the inverse transformation can be applied to obtain the proposed  $\mathbf{y}_i^*$ , sequentially for  $\ell = 1, \dots, d$ , where the bounds may also be updated sequentially if they depend on  $y_{i,1:\ell-1}^*$ , e.g. for age at first child in our application. This ordering also guarantees that the Jacobian matrix is lower triangular, so its determinant is simply the product of the diagonal elements,  $|\mathcal{J}_t(\mathbf{y}_i)| = \prod_{\ell=1}^d \mathcal{J}_{t,\ell,\ell}(y_{i,\ell}; \mathbf{y}_{i,1:\ell-1})$ , with  $\mathcal{J}_{t,\ell,\ell}(y_{i,\ell}; \mathbf{y}_{i,1:\ell-1}) = (u_{i,\ell} - l_{i,\ell}) / [(y_{i,\ell} - l_{i,\ell})(u_{i,\ell} - y_{i,\ell})]$ , for  $u_{i,\ell} \in \mathbb{R}$ ,  $l_{i,\ell} \in \mathbb{R}$ .

**SMC for adaptive truncation.** The second stage involves the selection of the truncation level  $J$  by sequentially increasing it from the initial level  $J_0$ . The addition of a new component improves the quality of the approximation to the infinite-dimensional model but increases the computational burden, due to the considerable number of parameters added. Therefore, devising an algorithm that can select the level of truncation parsimoniously is crucial. To achieve this, possible approaches are presented in Norets (2020) and Griffin (2016). We focus on the latter, which adaptively increases the number of mixture components via SMC.

The MCMC draws from the previous step are used as the  $M$  initial particles in the SMC. At each iteration of the SMC, a new component is added to the mixture, by sampling the additional set of parameters  $(w_{J+1}^m, \psi_{J+1}^m, \theta_{J+1}^m)$  from a suitable importance distribution. We sample from the prior  $\text{Beta}(v_{J+1}^m) \mathbf{P}_0(\psi_{J+1}^m, \theta_{J+1}^m)$ , independently for  $m = 1, \dots, M$ , making use of the recursive stick-breaking relation  $w_{J+1}^m = v_{J+1}^m [(1 - v_J^m)/v_J^m] w_J^m$ . The particle weights  $\tilde{\vartheta}_{J+1}^{1:M} = (\tilde{\vartheta}_{J+1}^1, \dots, \tilde{\vartheta}_{J+1}^M)$  are then updated as follows:

$$\tilde{\vartheta}_{J+1}^m = \tilde{\vartheta}_J^m \prod_{i=1}^n \frac{f_{\mathbf{P}_{\mathbf{x}_i}^{J+1}}(\mathbf{y}_i^m | \mathbf{w}_{1:J+1}^m, \psi_{1:J+1}^m, \theta_{1:J+1}^m)}{f_{\mathbf{P}_{\mathbf{x}_i}^J}(\mathbf{y}_i^m | \mathbf{w}_{1:J}^m, \psi_{1:J}^m, \theta_{1:J}^m)}.$$

When the effective sample size (ESS) is lower than a threshold, indicating poor mixing, the particle values are resampled according to such weights (Del Moral et al., 2006). Here, we resort to systematic resampling (Kitagawa, 1996) and perform a rejuvenating step (Gilks and Berzuini, 2001), where the particles are replaced with new values

sampled through  $m^*$  iterations of the adaptive MCMC with  $J_0 = J + 1$ . The SMC provides weighted samples from the sequence of truncated posteriors  $\mathbf{P}_J^n$ , converging to the infinite posterior  $\mathbf{P}^n$ . To decide when a sufficiently accurate approximation has been obtained, we follow Griffin (2016) and stop at the truncation level  $J^*$ , such that the discrepancy  $D(\mathbf{P}_J^n, \mathbf{P}_{J+1}^n) = |\text{ESS}_J - \text{ESS}_{J+1}|$  is less than a specified  $\delta > 0$ , for a fixed number  $I$  of consecutive increments,  $J = J^* - I + 1, \dots, J^*$ . We use the suggested values of  $\delta = 0.01M$ ,  $I = 4$ , and  $m^* = 3$ . As an alternative to the ESS, we also consider a discrepancy based on the conditional effective sample size (CESS), which was proposed by Zhou et al. (2016), in the context of model comparison via SMC.

**Simulation study.** To assess the performance of the proposed procedure, we applied our model to a simulated dataset with a known structure, mimicking the most relevant features of our motivating data. Specifically, we considered  $q^* = 3$  covariates; the first,  $x_1$ , is continuous and observed at a discrete scale (resembling the age at interview in our case study); the others,  $x_2^*$  and  $x_3^*$ , are categorical with three and two levels, respectively. We generate two positive integer-valued responses,  $Z_1$  and  $Z_2$ , and one binary response,  $Z_3$ , related in different extents to the covariates.  $Z_1$  is a discretized noisy observation of a nonlinear function of  $x_1$ . Similarly,  $Z_2$  is a discretized noisy observation of a nonlinear function of  $x_1$  and the realized  $z_1$ . In both cases, the response curves are the same for  $x_2^* = 2, 3$  and differ for other categorical combinations, while the errors are not normal but right skewed, additionally depending on  $x_1$  and  $x_3^*$  for the second response. Censoring is defined before discretization, when the responses are greater than the first covariate. Finally,  $Z_3$  was simulated from a linear probit model depending only on  $x_1$ . Complete details of the data-generating distributions are provided in the SM, together with the specification of the prior parameters.

We performed a robustness analysis on the simulated dataset, comparing several initializations, namely by setting  $J_0 = 2, 3, 5, 10, 15, 20, 30$ . We found that initializing the algorithm with a conservative number of components for moderate sample sizes provides a good compromise between computational time, mixing, and accuracy. For large sample sizes however, we suggest to initialize with a generous number of components because of the computational burden implied by resampling. Specifically, for modest sample sizes, we can save the parametric mixture likelihoods, unnormalized weights, and normalizing constants for each observation and for every particle, with a memory complexity of  $\mathcal{O}(nJM)$ . For large sample sizes, this becomes too costly and these terms need to be recomputed at each block update of the MCMC rejuvenation step of the SMC. Thus, it is convenient to minimize the need of resampling for large sample sizes, by not being too conservative in setting  $J_0$  for the problem at hand. We also considered two different discrepancy measures defining the stopping rule of the SMC, i.e. ESS and CESS, and results confirmed robustness to such choice.

Finally, focusing on the model flexibility and its ability to recover the correct structure present in the data, we performed additional simulations, exploring longer chains, increased sample size, and omitted censoring. The results, reported in the SM, are satisfactory for all scenarios, indicating (as expected) improvements when a longer chain is run or the censoring is removed.

Overall, our model was able to recover the underlying structure present in the data. The true conditional behavior was well recovered in areas where data is available. However, as can be expected, the model struggled when predicting at values far from the observed data. We highlight that interpretation of the conditional dependence structure in the latent scale as well as the latent covariance matrices of the mixture components and its relation to the dependence structure on the observed scale is an open and interesting direction of research, which would expand the work of [García-Zattera et al. \(2007\)](#) in the parametric setting. We also observed improvements in comparison with a parametric version of our model.

## 5 Application: life patterns of Colombian women

We employ our model to study the joint distribution of the ages at *Sexual Debut* ( $Z_1$ ), *Union* ( $Z_2$ ), and *First Child* ( $Z_3$ ) as well as *Work Status* ( $Z_4$ ) at the moment of the interview, conditional on the considered covariates. These are *Age* at interview ( $X_1$ ), *Region* ( $X_2^*$ ) and *Area* ( $X_3^*$ ) of residence, having (P) or not ( $\bar{P}$ ) been disciplined using *Physical Punishment* ( $X_4^*$ ) during childhood, and having (B) or not ( $\bar{B}$ ) been exposed to *Parental Domestic Violence* ( $X_5^*$ ), referring to whether the respondent witnessed her father beating her mother.

To specify our model, we define the link functions:

$$\begin{aligned} z_\ell &= h_\ell(\mathbf{y}, \mathbf{x}) = c_\ell(\mathbf{y}, \mathbf{x}) \lfloor \exp(y_\ell) \rfloor, \quad \text{for } \ell = 1, 2, \\ z_3 &= h_3(\mathbf{y}, \mathbf{x}) = c_3(\mathbf{y}, \mathbf{x}) \lfloor \exp(y_1) + \exp(y_3) \rfloor, \\ z_4 &= h_4(y_4, \mathbf{x}) = \mathbb{1}_{[0, \infty)}(y_4), \end{aligned}$$

with  $c_\ell(\mathbf{y}, \mathbf{x}) = \mathbb{1}_{(0, x_1+1)}(\exp(y_\ell))$ , for  $\ell = 1, 2$ , and  $c_3(\mathbf{y}, \mathbf{x}) = \mathbb{1}_{(0, x_1+1)}(\exp(y_1) + \exp(y_3))$ . In this case,  $\exp(y_1)$  and  $\exp(y_2)$  can be interpreted as the latent continuous ages at sexual debut and union, respectively. The constraint that age at first child must be greater than age at sexual debut is strictly enforced through the transformation, and we can interpret  $\exp(y_3)$  as the latent continuous time between sexual debut and first child and  $\exp(y_1) + \exp(y_3)$  as the latent continuous age at first child. The bounds required in the adaptive MCMC are obtained from inverting  $z_\ell = h_\ell(\mathbf{y}, \mathbf{x})$ ; concretely,

$$\begin{aligned} (l_\ell, u_\ell) &= \begin{cases} (\log(x_1 + 1), \infty) & \text{for censored } z_\ell = 0 \\ (\log(z_\ell), \log(z_\ell + 1)) & \text{for uncensored } z_\ell \neq 0 \end{cases}, \quad \text{when } \ell = 1, 2, \\ (l_3, u_3) &= \begin{cases} (\log(\max[0, x_{i,1} + 1 - \exp(y_{i,1})]), \infty) & \text{for } z_3 = 0 \\ (\log(\max[0, z_{i,\ell} - \exp(y_{i,1})]), \log(z_{i,\ell} - \exp(y_{i,1}) + 1)) & \text{for } z_3 \neq 0 \end{cases}, \\ (l_4, u_4) &= \begin{cases} (-\infty, 0) & \text{for } z_4 = 0 \\ (0, \infty) & \text{for } z_4 = 1 \end{cases}. \end{aligned}$$

The prior parameters are specified as follows. For the linear coefficients and covariance matrix of each component, they are set empirically based on a multivariate linear



regression fit. Specifically, we set  $y_{i,\ell} = (l_{i,\ell} + u_{i,\ell})/2$ . For  $\ell = 3$ , when the lower bound is  $-\infty$ , i.e. age at sexual debut is equal to age at first child, we set  $y_{i,3} = u_{i,3} - 1$ . For censored observations, we sample  $y_{i,\ell}$  from a truncated normal distribution with mean and covariance computed from the uncensored observations. For the binary response,  $y_{4,i} = -1$  for  $z_{4,i} = 0$  and  $y_{4,i} = 1$  for  $z_{4,i} = 1$ . A multivariate linear regression fit for this auxiliary response gives estimates  $\hat{\beta}$  of the linear coefficients and  $\hat{\Sigma}$  of the covariance matrix. We then define

$$\mathbb{E}[\beta_j] = \beta_0 = \hat{\beta} \quad \text{and} \quad \mathbb{E}[\Sigma_j] = \frac{1}{\nu - b - 1} \Sigma_0 = \hat{\Sigma}.$$

Together,  $\mathbf{U}$  and  $\Sigma_j$  reflect the variability of  $\beta_j$  across components, and we set the matrix  $\mathbf{U}$  such that  $\min(\text{diag}(\hat{\Sigma})) \mathbf{U} = 20(\mathbf{X}^\top \mathbf{X})^{-1}$ . The factor 20 for this g-prior was selected to ensure reasonable ages (i.e. mostly lower than 100) in prior simulations and to avoid very extreme age values that can result when the constant is too big (making the prior too vague). Indeed, we explored more uninformative and vague prior choices but found that this could lead to quite large and unreasonable imputed ages for censored data. We further set  $\nu = b + 3$ . Other specified hyperparameters include  $\mu_{0,1} = \bar{x}_1$ ;  $u_1 = 1/2$ ;  $\alpha_1 = 2$ ;  $\gamma_1 = u_1(\text{range}(x_{1:n,1})/4)^2$ ;  $\boldsymbol{\rho}_k = (1, 1)$  for  $k = 2, \dots, q$ ; and the parameters of the stick-breaking prior are  $\zeta_{j,1} = 1$  and  $\zeta_{j,2} = 1$ .

We initialize the MCMC algorithm with a number of components,  $J_0 = 35$ , large enough to avoid a small ESS and subsequent resampling. The MCMC is run for 20,000 iterations after discarding the first 30,000 as burn-in, and one in every 10 iterations is saved to produce 2,000 particles. For the SMC, we choose the ESS-based stopping rule, due to the robustness observed for the simulated data sets.

## 5.1 Posterior predictive checks

The assessment of the goodness of fit of the proposed model is of crucial importance when the data are of complex nature, such as the data analysed in this work. Therefore, before presenting the main results, we report posterior predictive checks in order to validate our model. Using the weighted particles produced by the algorithm in Section 4, we first replicate the data and use the replications to check the predictive power of our model. Figures 3 and 4 report such predictive checks for the regions Bogota and Territorios Nacionales. In particular, for the discretized ages at events, the figures compare the Kaplan-Meier curves for the data and each replicated data set. For the *Work* response, the histogram of the proportion of working women across the replicated datasets is displayed, along with the proportion in the observed data. In general, the estimates observed in the original data match the ones computed for the replicates, within all the considered categories, indicating good fit of the model to the data.

In addition, we compute the posterior predictive p-values (Gelman et al., 1996):

$$p(\mathbf{z}) = \mathbb{P}(T(\mathbf{z}^{rep}) \geq T(\mathbf{z}) | \mathbf{z}) = \int \mathbb{P}(T(\mathbf{z}^{rep}) \geq T(\mathbf{z}) | \mathbf{z}, \boldsymbol{\xi}) \pi(\boldsymbol{\xi} | \mathbf{z}) d\boldsymbol{\xi},$$

where  $T(\mathbf{z}^{rep})$  and  $T(\mathbf{z})$  represent the selected *discrepancy* computed respectively for replicated and observed data, and  $\boldsymbol{\xi} = (\mathbf{w}_{1:J^*}, \boldsymbol{\psi}_{1:J^*}, \boldsymbol{\theta}_{1:J^*})$  indicates the model parameters with posterior  $\pi(\boldsymbol{\xi} | \mathbf{z})$ . The integral above can be computed based on the weighted

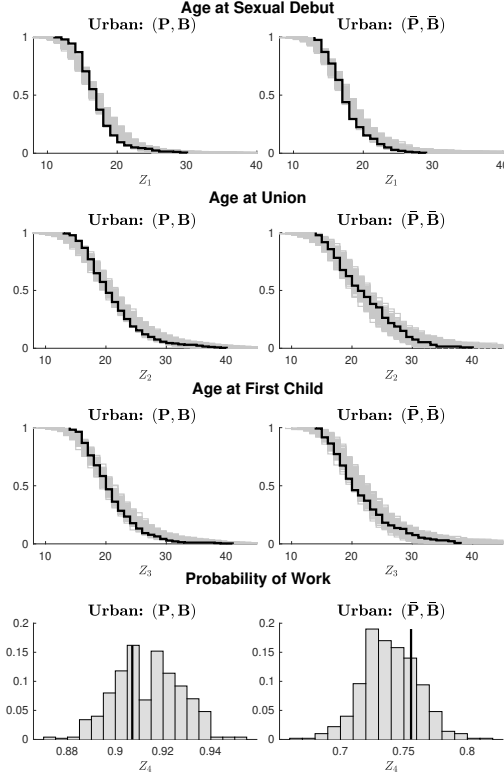


Figure 3: Predictive checks for the region Bogota, for women who grew up in violent  $(\mathbf{P}, \mathbf{B})$  and non-violent  $(\bar{\mathbf{P}}, \bar{\mathbf{B}})$  families. Rows (1-3): comparison between Kaplan-Meier curves of the discretized ages at events for the observed (black lines) and replicated (grey lines) data. Last row: comparison between proportions of working women in the observed (black vertical line) and replicated (grey histograms) data.

particles, by simulating a replicated dataset for each particle. In our application, the discrepancy measures used, for  $\ell = 1, 2, 3$ , are:

$$T^{\text{cens}}(\mathbf{c}_\ell) = \frac{1}{n} \sum_{i=1}^n |1 - c_{i,\ell} - \mathbb{P}(\tilde{Z}_\ell \geq (x_{i,1} + 1) | \mathbf{x}_i, \boldsymbol{\xi}^m)|,$$

$$T^{\text{noncens}}(\mathbf{z}_\ell) = \frac{1}{|A|} \sum_{i \in A} |z_{i,\ell} - \text{median}(\tilde{Z}_\ell; \boldsymbol{\xi}^m)|, \quad A = \{i \in \{1, \dots, n\} | c_{i,\ell} = 1\},$$

$$T(\mathbf{z}_4) = \frac{1}{n} \sum_{i=1}^n |z_{i,4} - \mathbb{P}(Z_4 = 1 | \mathbf{x}_i, \boldsymbol{\xi}^m)|.$$

Note that for the censored data, we only observe that the event is not experienced by the given age, and thus the discrepancy  $T^{\text{cens}}$  is defined as for a binary variable. On

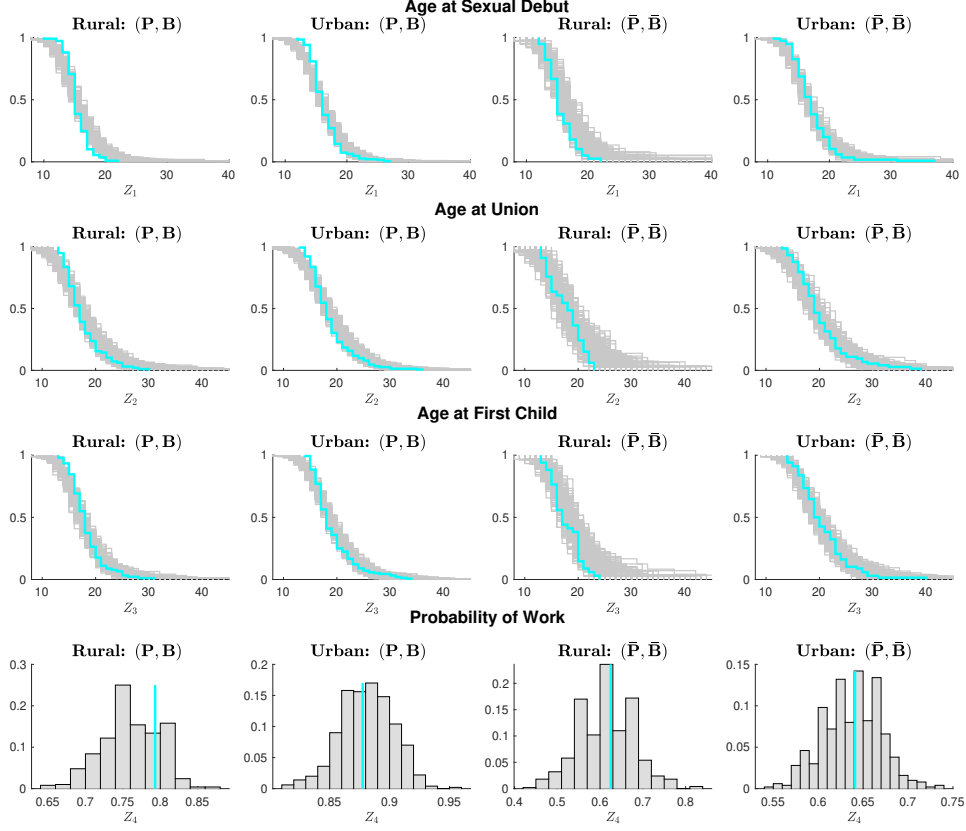


Figure 4: Predictive checks for the region Territorios Nacionales, in Rural or Urban areas, for women who grew up in violent ( $\mathbf{P}, \mathbf{B}$ ) and non-violent ( $\bar{\mathbf{P}}, \bar{\mathbf{B}}$ ) families. Rows (1-3): comparison between Kaplan-Meier estimators of the discretized ages at events for the observed (light blue lines) and replicated (grey lines) data. Last row: comparison between proportions of working women in the observed (light blue vertical line) and replicated (grey histograms) data.

the other hand, for non-censored events, the discrepancy  $T^{\text{noncens}}$  can be based on the estimated age at event. The posterior predictive p-values for the full sample and for subsets defined based on selected combinations of categorical covariates are reported in Table 2. Results are encouraging and support the goodness of fit observed in Figures 3 and 4, but highlight possible issues in interpreting censoring results.

## 5.2 Results

In the SM, we describe various posterior and predictive quantities that can be computed from the weighted particles to describe the relationship between the observed responses

		Full	Bogota		Terr. Nac.			
		Sample	Urban: ( $\mathbf{P}, \mathbf{B}$ )	Urban: ( $\bar{\mathbf{P}}, \bar{\mathbf{B}}$ )	Rural: ( $\mathbf{P}, \mathbf{B}$ )	Urban: ( $\mathbf{P}, \mathbf{B}$ )	Rural: ( $\bar{\mathbf{P}}, \bar{\mathbf{B}}$ )	Urban: ( $\bar{\mathbf{P}}, \bar{\mathbf{B}}$ )
$Z_1$	$c_i = 0$	0.148	0.032	0.885	0.792	0.989	0.697	0.826
	$c_i = 1$	1	1	1	1	1	0.878	0.994
$Z_2$	$c_i = 0$	0.056	0.227	0.843	0.9615	0.353	0.9945	0.725
	$c_i = 1$	1	0.991	0.658	1	1	0.966	0.931
$Z_3$	$c_i = 0$	0.315	0.591	0.124	0.671	0.765	0.998	0.269
	$c_i = 1$	0.939	0.986	0.611	0.982	0.3325	0.483	0.824
$Z_4$		0.604	0.363	0.935	0.872	0.547	0.775	0.604

Table 2: Posterior predictive p-values for selected discrepancies computed for the full sample and subsets corresponding to violent ( $\mathbf{P}, \mathbf{B}$ ) and non-violent ( $\bar{\mathbf{P}}, \bar{\mathbf{B}}$ ) families. We report the results obtained for the regions Bogota and Territorios Nacionales.

and the covariates. For the sake of conciseness, here we display only a selection of predictive quantities, which offer some insights about the situation of Colombian women. Specifically, we compare women who were raised in violent family environments ( $\mathbf{P}, \mathbf{B}$ ) with those who were not ( $\bar{\mathbf{P}}, \bar{\mathbf{B}}$ ). Figure 5 displays the predictive medians of the (undiscretized) ages at events and the posterior probability of working as functions of *Age*. More detailed information arises from the analysis of the predictive densities, some of which are reported in Figure 6. Notice that due to the clear asymmetry in the densities, the predictive median allows a better representation of the center, as opposed to the mean. We highlight that to improve visualization we focus on predictive quantities (medians, probability of success, densities, etc.), but the Bayesian approach permits to also study uncertainty and report credible intervals for these quantities (e.g. Figure 9).

The data presents heavy censoring for younger cohorts (summarized in Table 1). This information is included by imputing, at each iteration of the algorithm, ages at events which must be higher than *Age*. Indeed, above the dashed lines of Figure 6, the density estimates are based on these imputed ages and borrowing of information at other covariate levels. Therefore, while we can reliably estimate the mass above the dashed line given *Age*, caution should be used when interpreting the shape of the right tail in this region, as this is not identifiable from the observed data. Moreover, when such mass exceeds 0.5, the predictive median is affected by the imputed values and is therefore less reliable. This corresponds to median values of age at event which are higher than *Age*, represented as dotted lines in the figures. Further, censored data also arise from women who will never experience an event. This is the prevailing cause of censoring for the older cohorts, contributing to higher medians and heavier right tails. Accommodating censored cases is clearly useful; however, results arising from heavily censored data should be interpreted with caution.

Starting with Figure 5, observe that the shapes of the median curves change across combinations of the categorical covariates, which justifies the employment of a flexible model that does not impose a single functional form. A clear difference is evident between urban and rural areas, the latter presenting lower ages at events, controlling for other covariates. This is expected since rural areas are generally characterized by lower levels of education and wealth indicators, both identified in the literature as factors

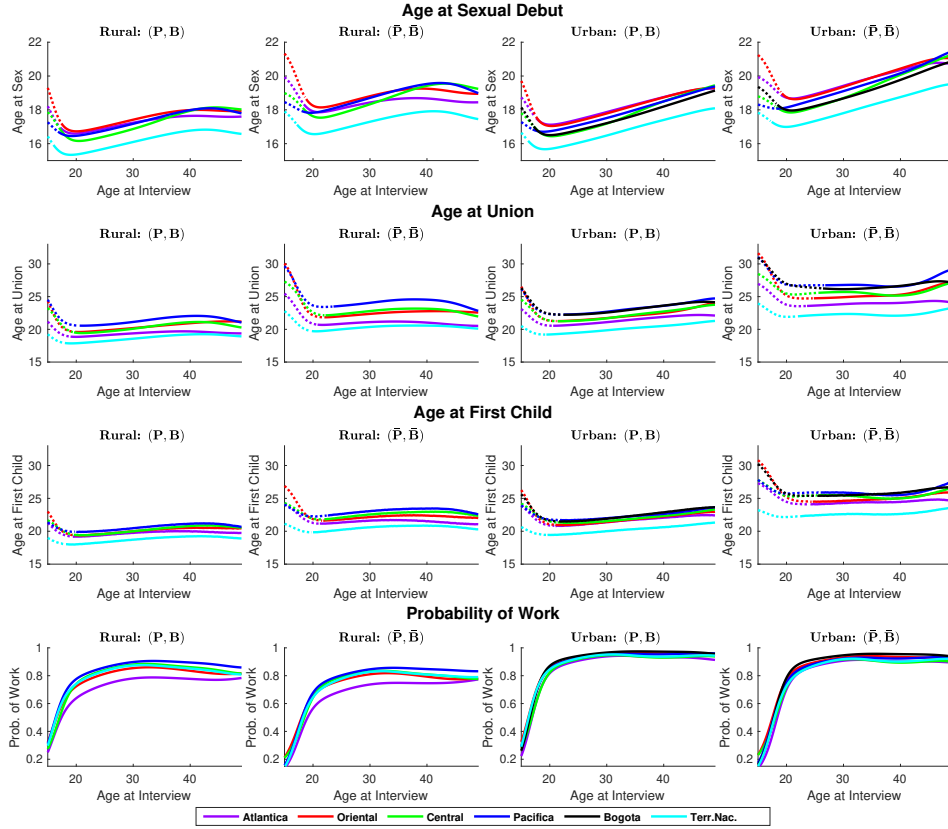


Figure 5: Predictive medians of the ages at sexual debut, union and first child, and posterior probability of working, as functions of *Age*, for women who grew up in violent ( $P, B$ ) and non-violent ( $\bar{P}, \bar{B}$ ) families. Dotted lines indicate when the median exceeds *Age*.

related to anticipation of sexual activity and family formation. Comparing cohorts, we observe that younger women tend to anticipate sexual debut, a phenomenon largely recognized as a consequence of the better knowledge and the more diffuse use of contraceptive methods. Instead, the curves for the ages at union and at first child appear flatter, particularly for urban women with non-violent family environments and are even increasing for women from violent families. At first, this may seem counter-intuitive, because one would expect the younger generations to postpone family formation, particularly in urban areas, due to an expected prolonged education. However, an incorrect use of contraceptive methods, particularly among very young or less educated women, and the violent conditions linked to the armed conflict may result in early pregnancies (Ali et al., 2003; Núñez and Flórez, 2001; Daniels, 2015). Indeed, an increase in teenage childbearing in Colombia has been observed since 1990, mainly among women from

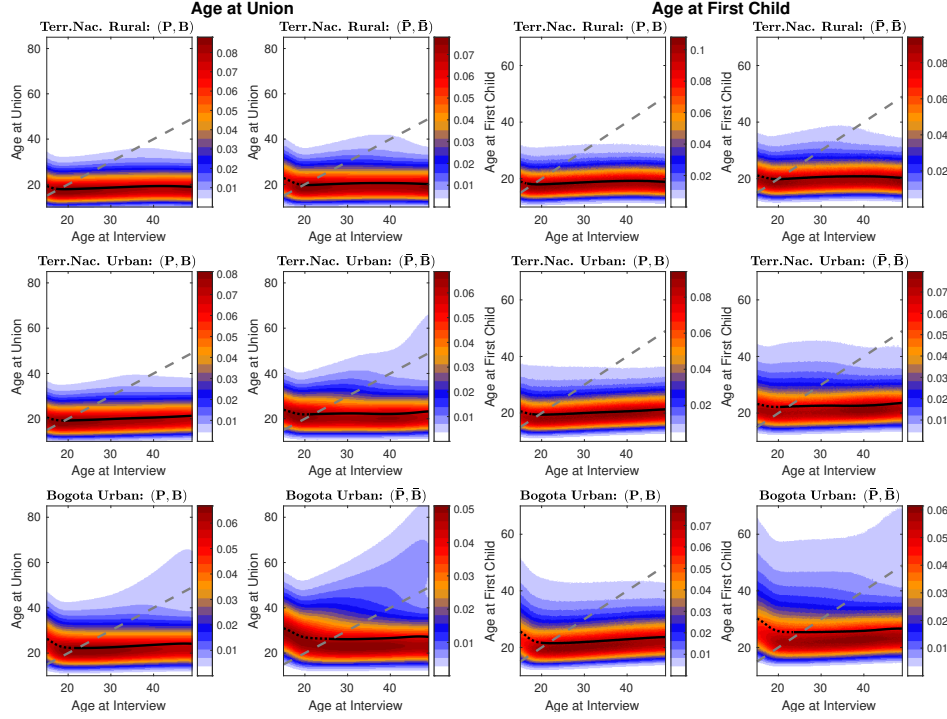


Figure 6: Predictive densities of the ages at union and first child as functions of *Age* for women who grew up in violent ( $\mathbf{P}, \mathbf{B}$ ) and non-violent ( $\bar{\mathbf{P}}, \bar{\mathbf{B}}$ ) families. Results are reported for urban and rural areas of the least developed region (Territorios nacionales) and for the capital (Bogota). The region above the dashed line indicates when age at event exceeds *Age*. The black line is the posterior median function.

disadvantaged backgrounds (Batyra, 2016; Flórez and Soto, 2007, 2013).

Focusing on the predictive densities for the least and the most developed regions, Territorios Nacionales and the capital city Bogota (Figure 6), further justifies the use of a density regression model. In fact, the observed flat median curves correspond to rather different distributional behaviors of ages at union and child, across covariate values. Moving from the least to the most developed context (top to bottom in the figure) entails an increase of the median curves, dispersion, and probabilities of not having experienced the events by a given *Age*. An increased dispersion, with pronounced right-skewness, is more evident for older cohorts in urban environments. This is in line with the greater heterogeneity in urban contexts as well as with the wider range of opportunities offered, for example in terms of education. Such heterogeneity becomes more pronounced among the older cohorts who have had time to profit from such opportunities. The flexibility gained in urban contexts is offset in violent environments, thus resulting in more concentrated distributions. While our definition of a violent environ-

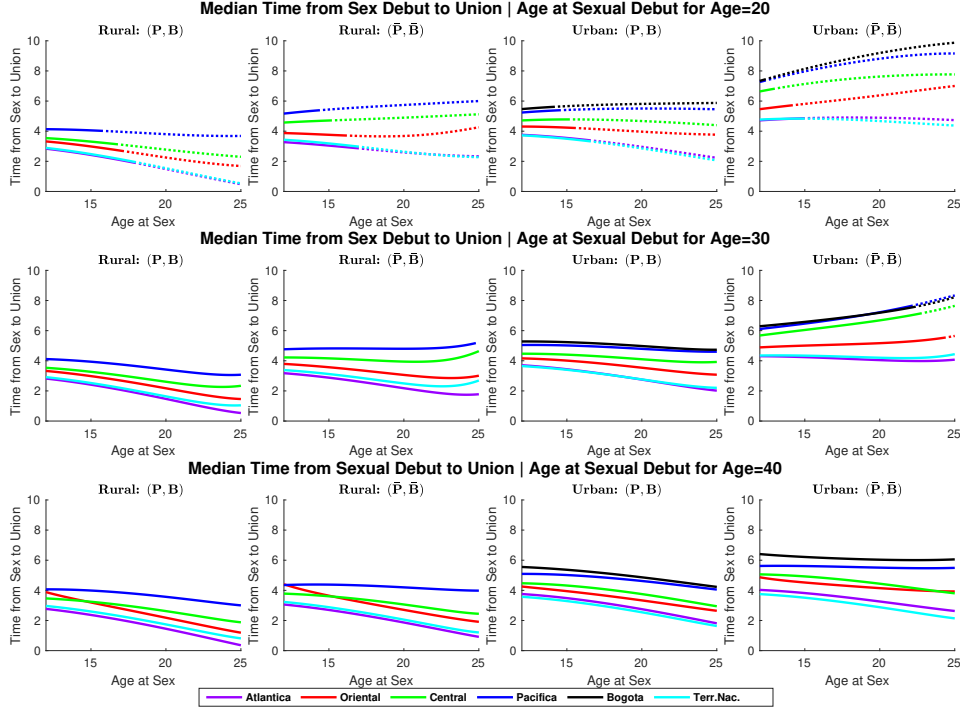


Figure 7: Predictive medians of the time from sexual debut to union conditional on age at sexual debut, as a function of the latter, for women with  $Age = 20, 30, 40$ , who grew up in violent ( $P, B$ ) and non-violent families ( $\bar{P}, \bar{B}$ ). Dotted lines indicate ages at event higher than the  $Age$ .

ment is not formal and refers only to the adoption of physical punishment methods and exposure to parental violence, the results signal the detrimental effect of family violence on Colombian women life patterns, and provides quantitative support for psychosocial studies (Restrepo Martínez et al., 2017).

The joint modeling approach permits us to study also the conditional relation between responses. For example, Figure 7 shows the conditional predictive medians of the time from sexual debut to union given the age at sexual debut for women with  $Age = 20, 30, 40$  (dotted lines indicate predicted ages at event higher than  $Age$ ; the corresponding conditional densities are reported in the SM). Interesting differences can be observed across regions, likely related to their socio-demographic characteristics (Ojeda et al., 2011). We observe that women in Atlantica and Territorios Nacionales (and to a lesser extent Oriental) compared with Pacifica and Bogota tend to experience sexual debut and union closer in time, suggesting that sexual debut is possibly delayed until union. Such tendency is more pronounced, compared to the other regions, for rural women raised in violent families. Similar results are observed for the time from sexual



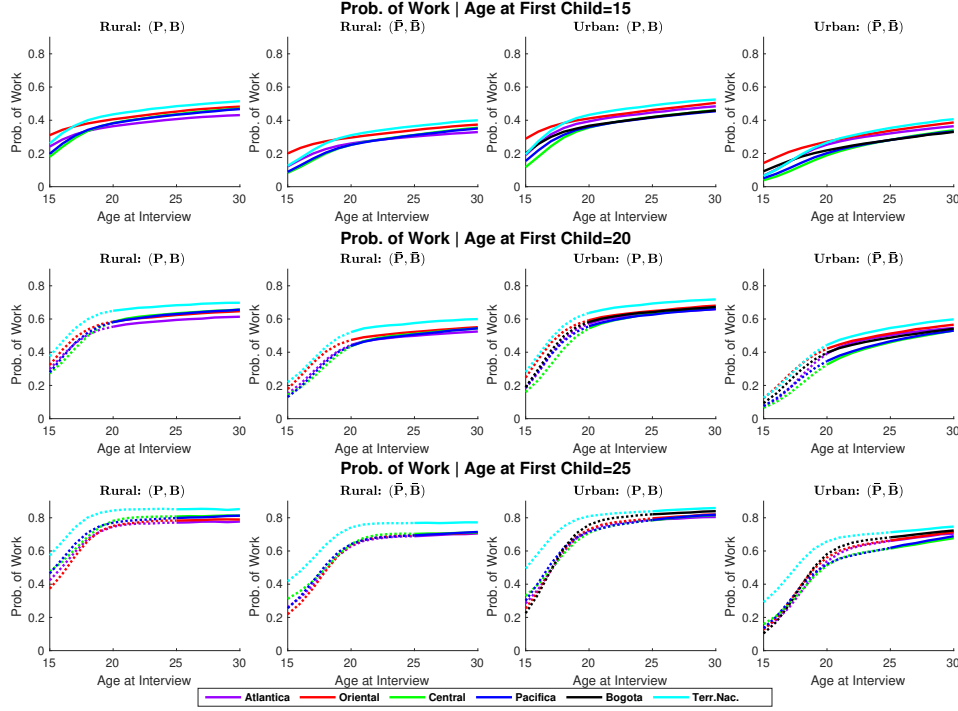


Figure 8: Predictive probability of working as function of *Age* conditional on different ages at first child, for women who grew up in violent ( $\mathbf{P}, \mathbf{B}$ ) and non-violent ( $\bar{\mathbf{P}}, \bar{\mathbf{B}}$ ) families. Dotted lines indicate ages at event higher than *Age*.

debut to child (details in the SM).

Finally, the probability of working (Figure 5, bottom row) is, as expected, higher in urban areas. Moreover, women who grew up in violent environments show a higher propensity to work, more pronounced among younger women. These same women, as previously observed, show a tendency to anticipate events. A possible explanation is that young women who leave the parental house to escape violence may start cohabitation and decide to drop out of school, entering the labor market to contribute to family income. This apparently contradicts studies (see e.g. [Gimenez Duarte et al., 2015](#)) pointing to the difficulties of young women, especially those with children, to participate in the labor market. However, this paradox is solved when analyzing the estimated predictive probabilities of working as functions of *Age*, conditional on having the first child at ages 15, 20 and 25 (Figure 8, top to bottom). Indeed, the probability of working at each *Age* increases with the age at first child. In particular, we observe a much lower probability of working for young mothers, that persists even when considering their labor market participation later in life. This suggests a scaring effect of teenage motherhood. This effect is further supported by Figure 9, which depicts the 95% point-

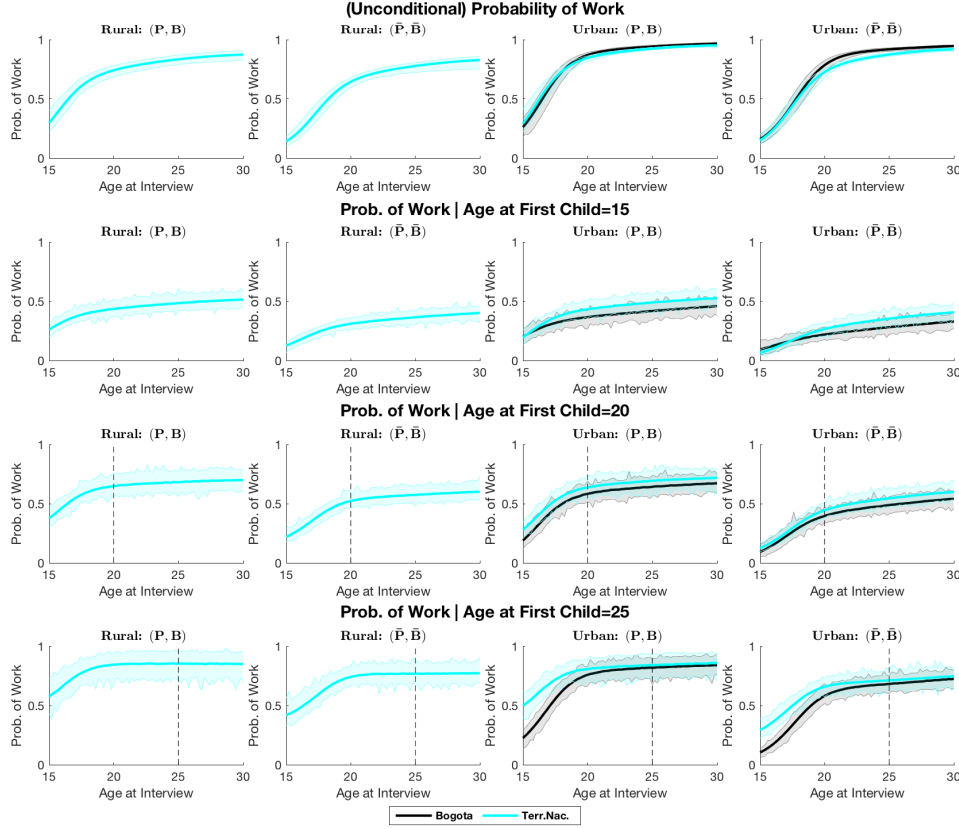


Figure 9: Predictive (unconditional) probability of working as function of *Age* with 95% pointwise credible intervals (top row) for women who grew up in violent ( $\mathbf{P}, \mathbf{B}$ ) and non-violent ( $\bar{\mathbf{P}}, \bar{\mathbf{B}}$ ) families. Subsequent rows depict the probability conditional on different ages at first child. Results are reported for urban and rural areas of the least developed region (Territorios Nacionales) and for the capital (Bogota). Left of the dashed line indicates when the age at first child is higher than *Age*.

wise credible intervals for the probability of working as functions of *Age*, comparing the unconditional probability (top row) and the probability conditional on different ages at first child (subsequent rows). Consistent with previous results, we focus on the least developed region (Territorios Nacionales) and on the capital (Bogota).

## 6 Concluding remarks

In this work, we proposed a novel Bayesian nonparametric model for density regression, allowing for mixed responses with censored, constrained, and binary traits, that

can flexibly change with combinations of the categorical and numerical covariates. We developed a general algorithm for posterior inference, that effectively scales to large datasets by adaptively determining the necessary truncation level to approximate the infinite-dimensional posterior. We customized the model and algorithm to a specific case study, but they can be applied in other contexts through minor modifications, by appropriate definition of the link functions. Note that our model accommodates for non-informative (or random) censoring. Interesting extensions concern other types of censoring. From a technical point of view, our results highlight the advantage of a flexible model, accounting for a different shape, location, and dispersion of the response distribution across the covariate levels, as well as for censoring. Additionally, a variety of classic graphic tools and quantities of interest, such as survival curves and hazard functions, can be derived. Importantly, the joint analysis of the responses allows for a rich variety of conditional analyses, which can be conducted focusing on different aspects, a very useful feature when studying complex phenomena.

For our case study, the findings suggest interesting considerations regarding life patterns of Colombian women. In the first instance, we found a confirmation of the differences between rural and urban areas, which evidence the need of interventions towards a more balanced development of the country. Furthermore, our results signal that the regions with a higher risk of early transition to adulthood are those with the worse development and wellness indicators, thus corroborating studies on the risks related to disadvantageous conditions. One of the most interesting results is the rather clear evidence of the impact of family violence on women’s choices and behaviors. An anticipation of the considered events is observed for women who were physically punished during childhood and witnessed parental domestic violence, two factors we used as proxies for a violent family environment. The relation between child abuse and neglect and the child’s future family choices has been discussed in the literature. Nonetheless, to our knowledge, this is the first attempt to study the possible relation between parental family violence and the events marking the transition to adulthood. Our findings confirm that a violent family environment can be regarded as a key risk factor that may nullify the positive influence of developed areas.

Overall, our case study may contribute to the planning of targeted interventions. Even if recent governments have shown an increased attention to the conditions of women and children, a formal statistical approach to systematically identify and quantify critical situations is crucial to support such a process. For example, teenage pregnancy is recognized as a priority issue in Colombia by the Government ([Gimenez Duarte et al., 2015](#); [Daniels, 2015](#)), due to its hindering personal development and agency ([Azevedo et al., 2012](#)); our results confirm its scaring effect and quantify the risk of teenage pregnancy, identifying some of the most vulnerable groups. We conclude with the hope that the present work may stimulate further reflection, research and survey on the topic, and possibly lead to additional investigations exploiting the availability of DHS surveys on other developing countries and the flexibility and wide applicability of our model.

## References

- Ali, M. M., Cleland, J., and Shah, I. H. (2003). “Trends in reproductive behaviour among young single women in Colombia and Peru: 1985-1999.” *Demography*, 40: 659–673.
- Antoniano-Villalobos, I., Wade, S., and Walker, S. G. (2014). “A Bayesian nonparametric regression model with normalized weights: A study of hippocampal atrophy in Alzheimer’s disease.” *Journal of the American Statistical Association*, 109(506): 477–490.
- Azevedo, J. P., Favara, M., Haddock, S. E., Lopez-Calva, L. F., Müller, M., and Perova, E. (2012). *Teenage pregnancy and opportunities in Latin America and the Caribbean*. Washington, D.C.: World Bank Group.
- Bao, J. and Hanson, T. E. (2015). “Bayesian nonparametric multivariate ordinal regression.” *Canadian Journal of Statistics*, 43(3): 337–357.
- Barrientos, A. F., Jara, A., and Quintana, F. A. (2012). “On the support of MacEachern’s dependent Dirichlet processes and extensions.” *Bayesian Analysis*, 7(2): 277–310.
- (2017). “Fully nonparametric regression for bounded data using dependent Bernstein polynomials.” *Journal of the American Statistical Association*, 112(518): 806–825.
- Batyra, E. (2016). “Fertility and the changing pattern of the timing of childbearing in Colombia.” *Demographic Research*, 35: 1343–1372.
- Canale, A. and Dunson, D. B. (2011). “Bayesian kernel mixtures for counts.” *Journal of the American Statistical Association*, 106: 1528–1539.
- Daniels, J. P. (2015). “Tackling teenage pregnancy in Colombia.” *The Lancet*, 385(9977).
- De Iorio, M., Müller, P., Rosner, G. L., and MacEachern, S. N. (2004). “An ANOVA model for dependent random measures.” *Journal of the American Statistical Association*, 99: 2205–215.
- De Yoreo, M. and Kottas, A. (2018a). “Bayesian nonparametric modeling for multivariate ordinal regression.” *Journal of Computational and Graphical Statistics*, 27: 71–84.
- (2018b). “Modeling for dynamic ordinal regression relationships: An application to estimating maturity of rockfish in California.” *Journal of the American Statistical Association*, 113(521): 68–80.
- De Yoreo, M., Reiter, J. P., and Hillygus, D. S. (2017). “Bayesian mixture models with focused clustering for mixed ordinal and nominal data.” *Bayesian Analysis*, 12: 679–703.
- Del Moral, P., Doucet, A., and Jasra, A. (2006). “Sequential Monte Carlo samplers.” *Journal of the Royal Statistical Society: Series B*, 68(3): 411–436.

- Dunson, D. B. and Park, J. H. (2008). “Kernel stick-breaking processes.” *Biometrika*, 95: 307–323.
- Flórez, C. E. and Soto, V. E. (2007). “Fecundidad adolescente y desigualdad en Colombia.” *Notas de Población*, 83: 41–74.
- (2013). “Factores protectores y de riesgo del embarazo adolescente en Colombia.” *Estudios a profundidad 2010. Encuesta nacional de demografía y salud (ENDS 1990/2010)*, Bogotá: Profamilia.
- García-Zattera, M. J., Jara, A., Lesaffre, E., and Declerck, D. (2007). “Conditional independence of multivariate binary data with an application in caries research.” *Computational Statistics & Data Analysis*, 51(6): 3223–3234.
- Gelman, A., Meng, X.-L., and Stern, H. (1996). “Posterior predictive assessment of model fitness via realized discrepancies.” *Statistica Sinica*, 733–760.
- Ghosh, J. K. and Ramamoorthi, R. V. (2003). *Bayesian Nonparametrics*. Springer-Verlag New York.
- Gilks, W. R. and Berzuini, C. (2001). “Following a moving target-Monte Carlo inference for dynamic Bayesian models.” *Journal of the Royal Statistical Society: Series B*, 63(1): 127–146.
- Gimenez Duarte, L. R., Van Wie, S. H., Muller, M., Schutte, R. F., Rounseville, M. Z., and Viveros Mendoza, M. C. (2015). “Enhancing youth skills and economic opportunities to reduce teenage pregnancy in Colombia (English).” World Bank Report 97822-CO, Washington, D.C.: World Bank Group.
- Griffin, J. E. (2016). “An adaptive truncation method for inference in Bayesian non-parametric models.” *Statistics and Computing*, 26: 423–441.
- Griffin, J. E. and Steel, M. (2006). “Order-based dependent Dirichlet processes.” *Journal of the American Statistical Association*, 10: 179–194.
- Griffin, J. E. and Stephens, D. A. (2013). “Advances in Markov chain Monte Carlo.” In *Bayesian Theory and Applications*. Oxford University Press.
- Hanson, T. and Johnson, W. O. (2004). “A Bayesian semiparametric AFT model for interval-censored data.” *Journal of Computational and Graphical Statistics*, 13(2): 341–361.
- Ishwaran, H. and James, L. F. (2001). “Gibbs sampling methods for stick-breaking priors.” *Journal of the American Statistical Association*, 96: 161–173.
- Jara, A., García-Zattera, M. J., and Lesaffre, E. (2007). “A Dirichlet process mixture model for the analysis of correlated binary responses.” *Computational Statistics & Data Analysis*, 51(11): 5402–5415.
- Jara, A., Lesaffre, E., De Iorio, M., and Quintana, F. (2010). “Bayesian semiparametric inference for multivariate doubly-interval-censored data.” *Annals of Applied Statistics*, 4(4): 2126–2149.

- Kitagawa, G. (1996). “Monte Carlo filter and smoother for non-Gaussian nonlinear state space models.” *Journal of Computational and Graphical Statistics*, 5(1): 1–25.
- Korsgaard, I. R., Lund, M. S., Sorensen, D., Gianola, D., Madsen, P., and Jensen, J. (2003). “Multivariate Bayesian analysis of Gaussian, right censored Gaussian, ordered categorical and binary traits using Gibbs sampling.” *Genetics Selection Evolution*, 35(2): 159–183.
- Kottas, A. and Gelfand, A. E. (2001). “Bayesian semiparametric median regression modeling.” *Journal of the American Statistical Association*, 96(456): 1458–1468.
- Kottas, A., Müller, P., and Quintana, F. A. (2005). “Nonparametric Bayesian modeling for multivariate ordinal data.” *Journal of Computational and Graphical Statistics*, 14(3): 610–625.
- Lo, A. Y. (1984). “On a class of Bayesian nonparametric estimates: I. Density estimates.” *Annals of Statistics*, 351–357.
- MacEachern, S. N. (1999). “Dependent nonparametric processes.” In *ASA Proceedings of the Section on Bayesian Statistical Science*, 50–55. Alexandria, VA: American Statistical Association.
- (2000). “Dependent Dirichlet processes.” Technical report, Department of Statistics, Ohio State University.
- Müller, P., Erkanli, A., and West, M. (1996). “Bayesian curve fitting using multivariate normal mixtures.” *Biometrika*, 88: 67–79.
- Norets, A. (2020). “Optimal auxiliary priors and reversible jump proposals for a class of variable dimension models.” *Econometric Theory*.
- Norets, A. and Pelenis, J. (2020). “Adaptive Bayesian estimation of mixed discrete-continuous distributions under smoothness and sparsity.” *Journal of Econometrics*.
- Núñez, J. and Flórez, C. E. (2001). “Teenage Childbearing in Latin American Countries.” IDB Working Paper No. 147, Inter-American Development Bank, Research Department. Available at <http://dx.doi.org/10.2139/ssrn.1814694>.
- Ojeda, G., Ordóñez, M., and Ochoa, L. H. (2011). *Colombia Encuesta Nacional de Demografía y Salud 2010*. Bogotá, Colombia: Profamilia. Available at <http://dhsprogram.com/pubs/pdf/FR246/FR246.pdf>.
- Papageorgiou, G., Richardson, S., and Best, N. (2015). “Bayesian non-parametric models for spatially indexed data of mixed type.” *Journal of the Royal Statistical Society: Series B*, 77(5): 973–999.
- Pati, D., Dunson, D. B., and Tokdar, S. T. (2013). “Posterior consistency in conditional distribution estimation.” *Journal of Multivariate Analysis*, 116: 456 – 472.
- Restrepo Martínez, M., Trujillo Numa, L., Restrepo Bernal, D., Torres de Galvis, Y., and Sierra, G. (2017). “Sexual abuse and neglect situations as risk factors for adolescent pregnancy.” *Revista Colombiana de Psiquiatría (English ed.)*, 46(2): 74–81.

- Roberts, G. O., Gelman, A., and Gilks, W. R. (1997). “Weak convergence and optimal scaling of random walk Metropolis algorithms.” *Annals of Applied Probability*, 7: 110–120.
- Roberts, G. O. and Rosenthal, J. S. (2001). “Optimal scaling for various Metropolis-Hastings algorithms.” *Statistical Science*, 16: 351–367.
- Rodriguez, A. and Dunson, D. B. (2011). “Nonparametric Bayesian models through probit stick-breaking processes.” *Bayesian Analysis*, 6: 145–178.
- Shahbaba, B. and Neal, R. (2009). “Nonlinear models using Dirichlet process mixtures.” *Journal of Machine Learning Research*, 10: 1829–1850.
- Wade, S., Dunson, D. B., Petrone, S., and Trippa, L. (2014a). “Improving prediction from Dirichlet process mixtures via enrichment.” *Journal of Machine Learning Research*, 15: 1041–1071.
- Wade, S., Walker, S. G., and Petrone, S. (2014b). “A predictive study of Dirichlet process mixture models for curve fitting.” *Scandinavian Journal of Statistics*, 41(3): 580–605.
- Zhou, Y., Johansen, A. M., and Aston, J. A. D. (2016). “Toward automatic model comparison: an adaptive sequential Monte Carlo approach.” *Journal of Computational and Graphical Statistics*, 25(3): 701–726.

### **Acknowledgments**

The work reported in this paper was funded by the University of Warwick Academic Returners Fellowship and the University of Oslo. Raffaella Piccarreta acknowledges MIUR-bando PRIN 2017 for the support in her contribution to the final article.

The authors are grateful to the Editor, the Associate Editor and the referees for their thoughtful and helpful comments on a previous version of the manuscript.

See the Supplementary Material for further details. The code is publicly available, along with the simulated data to reproduce results:

[https://github.com/sarawade/BNPDensityRegression\\_AdaptiveTruncation](https://github.com/sarawade/BNPDensityRegression_AdaptiveTruncation)